SYSTEM AND METHOD FOR AUTOMATED OPTIMIZATION
OF SEARCH RESULT RELEVANCE

CROSS-REFERENCE TO RELATED APPLICATION

5   This application claims the benefit of U.S. Provisional Application No. 60/535,353, filed January 9, 2004, which is hereby claimed under 35 U.S.C. § 119.

FIELD OF THE INVENTION

In general, the present invention relates to computer software and search engines and, in particular, to systems and methods for automating search result relevance optimization.

10   ## BACKGROUND OF THE INVENTION

The Internet search engine has become an important source of revenue for the service providers that operate them. The revenue is primarily generated from the display of advertisements to search engine users. The more Internet traffic that a search engine receives, the more attractive it is to advertisers and the more revenue it can generate. It is

15   generally regarded that the best way search engines can increase traffic is to provide highly relevant search results. But what is relevant today may not be relevant tomorrow or even relevant later the same day. It is difficult for service providers to keep pace with the rapid changes in searchable content based on seasonal and popular trends and topical events in the news.

20   One way that search engine operators strive to maintain the relevance of the results that their search engines generate is to use a relevance schema. The relevance schema represents the algorithm the search engine uses to generate a set of search results, usually in a

particular order of relevance. The relevance schema is continually reevaluated using human judges to determine whether the results produced using the schema are valid, i.e., whether the results are still relevant. The search engine operator makes changes to the schema from time to time, as indicated by the human judges.

5   The problem with the above approach to maintaining the relevance of the search engine is that it is time-consuming, slow, and subjective. The human judges can only evaluate just so many possible search results, and their judgments of what is or is not relevant may not reflect a typical user's judgment. Other approaches suffer from similar drawbacks. For example, some users may respond to surveys conducted by the search

10   engine operator, giving direct feedback on the relevance of a particular set of search results. But the amount of data collected in this manner may be of insufficient volume to be considered reliable, and simply does not have the breadth and scale to truly reflect what users want when conducting their searches.

Another approach that is becoming more prevalent is the use of click-through data

15   collected for the search results. The search engine operator collects the user's interaction with the search results by recording the number of times users click on a result, referred to as the "click-through rate" or CTR. The click-through data has a number of advantages in that data can be collected in large volume as users interact with search results and is therefore a more objective measure of user satisfaction and more reliable predictor of relevance. In

20   general, experience has shown that the higher the CTR, the more relevant the result, or at least the greater the satisfaction of the user with the result. But the CTR data must still be analyzed and the operator must then decide how to update the relevance schema to generate better results. Moreover, the CTR data alone may be insufficient to produce a meaningful result. For example, the CTR of a particular result may be influenced by a number of factors

25   related to the appearance of the result on the page that cause the CTR to be unduly inflated out of proportion to the actual relevance of the underlying result.

No matter what the approach, determining the relevance of search results is a difficult task, in large part because there is no single definitive indicator of success of a search result. The sheer scale of the number of queries handled by a search engine and the speed with

30   which the search results are generated make relevance a fast-moving target.

SUMMARY OF THE INVENTION

To overcome the above-described problems, a system, method, and computer-accessible medium for automating the optimization of search result relevance in a search engine are provided. The system and method continually collect data that represent various aspects of how a search result is performing and compare that performance data to the expected performance for the search result. The system and method further diagnose the possible causes of underperforming results and automatically adjust the search engine operation to optimize the search result relevance.

In accordance with one aspect of the present invention, the performance data is collected from one or more sources and preferably includes implicit data automatically collected when the user interacts with the search result, such as CTR data, but may also include explicit data collected when the user makes use of search engine help or support features, or responds to user satisfaction surveys, as well as subjective, human-judged data, relevance verification test data, and sample test data. The various sources of data may be normalized to reflect their relative importance or reliability as predictors of relevance.

In accordance with another aspect of the present invention, the expected performance data may be one or more value(s) that represents the expected performance for a result, such as an expected CTR, and can vary from one market to the next. A result may be determined to be underperforming when its performance does not meet the expected performance, including performing below or substantially below the expected performance value.

In accordance with a further aspect of the present invention, diagnosing the possible causes of underperforming results includes considering a number of factors, such as whether the result links to a Web site or document that is no longer valid, is appearing in a poor location, whether the search term that produced the result is easily misspelled or too broad to produce a meaningful result, or whether the search for a particular search term should be constrained to a particular resource, such as a local community resource.

In accordance with a still further aspect of the present invention, automatically adjusting the search engine operation to optimize the search result relevance includes taking a variety of actions on components of the search engine. For example, the search schema used by the search engine to produce the search results may be modified so that future search

results may be reranked, removed, repositioned, or replaced with different results. In some cases, the search results may remain the same, but the presentation of the results may be augmented to increase visibility and improve performance. In still other cases, the underperforming search may be temporarily modified to include new or augmented results

5 and tested in a sample market before making the modification permanent. In cases where the search term is itself the problem, the spellchecker tolerance may be increased, or the presentation of the results may be modified to prompt the user to clarify or narrow the search with additional search terms. In some cases, the operation of the search engine may be adjusted in real time to rapidly optimize the relevance of the search results in response to

10 sudden changes in search result performance.

In accordance with yet other aspects of the present invention, a computer-accessible medium for automating the optimization of search result relevance in a search engine is provided. The computer-accessible medium comprises data structures and computer-executable components comprising an automated relevance optimizer for collecting

15 performance data, diagnosing underperforming searches, and automatically adjusting the operation of the search engine to optimize the relevance of search results. The data structures define search result and performance data in a manner that is generally consistent with the above-described method. Likewise, the computer-executable components are capable of performing actions generally consistent with the above-described method.

20 BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same become better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

25 FIGURE 1 is a block diagram depicting an exemplary search engine system and one suitable operating environment in which the optimization of search result relevance may be automated, in accordance with the present invention;

FIGURE 2 is a block diagram depicting an automated relevance optimization system for implementing an embodiment of the present invention;

FIGURE 3 is a block diagram depicting in further detail an arrangement of certain components of the automated relevance optimization system of FIGURE 2 implemented in a search engine server of FIGURE 1, in accordance with an embodiment of the present invention;

FIGURE 4A is a pictorial diagram of an exemplary search engine user interface for implementing an embodiment of the present invention;

FIGURE 4B is a pictorial diagram of the exemplary search engine user interface of FIGURE 4A at a later time, after the search results have been automatically optimized in accordance with an embodiment of the present invention; and

FIGURES 5A-5B are flow diagrams illustrating the logic performed in conjunction with the automated relevance optimization system of FIGURES 2 and 3 for automating the optimization of search result relevance in a search engine in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

The following discussion is intended to provide a brief, general description of a computing system suitable for implementing various features of an embodiment of the invention. While the computing system will be described in the general context of a personal and server computer or other types of computing devices usable in a distributed computing environment where complementary tasks are performed by remote computing devices linked together through a communication network, those skilled in the art will appreciate that the invention may be practiced with many other computer system configurations, including multiprocessor systems, minicomputers, mainframe computers, and the like. In addition to the more conventional computer systems described above, those skilled in the art will recognize that the invention may be practiced on other computing devices including laptop computers, tablet computers, personal digital assistants (PDAs), cellular telephones, and other devices upon which computer software or other digital content is installed.

While aspects of the invention may be described in terms of programs or processes executed by a Web browser in conjunction with a personal computer or programs or processes executed by a search engine in conjunction with a server computer, those skilled in the art will recognize that those aspects also may be implemented in combination with other

program modules. Generally, program modules include routines, subroutines, programs, processes, components, data structures, functions, interfaces, objects, etc., which perform particular tasks or implement particular abstract data types.

FIGURE 1 is a depiction of an exemplary search engine system 100 and one suitable operating environment in which the optimization of search results may be automated in accordance with an embodiment of the present invention. As shown, the operating environment includes a search engine server 112 that is generally responsible for providing front-end user communication with various user devices, such as devices 102 and 104, and back-end searching services. The front-end communication provided by the search engine server 112 may include, among other services, generating text and/or graphics organized as a search Web page 106 using hypertext transfer protocols in response to information and search queries received from the various user devices, such as a computer system 102 and a personal digital assistant (PDA) 104. The back-end searching services provided by the search engine server 112 may include, among other services, using the information and search queries received from the various user devices 102, 104 to search for relevant Web content, generating search results 110 representing links to relevant Web content on the search Web page 106, and tracking Web page and search result performance.

In the environment shown in FIGURE 1, the search engine server 112 generates a search Web page 106 into which a user may input search terms 108 to initiate a search query for Web content 118 via the Internet 116. The search terms 108 are transmitted to a search engine server 112, which uses the terms to perform a search for Web content 120 that is relevant to the search terms 108 in accordance with a relevance schema 114. The relevance schema 114 is an algorithm that is periodically or continuously updated for use by the search engine to produce the most relevant results possible for any given search. The search engine server 112 relays the relevant Web content as a set of search results 110 for display to the user in the search Web page 106.

In the environment shown in FIGURE 1, the user devices 102, 104 communicate with a search engine server 112 via one or more computer networks, such as the Internet 116. Protocols and components for communicating via the Internet are well known to those of ordinary skill in the art of computer network communications. Communication between user

devices 102, 104, the search engine server 112, and the relevance schema 114 may also be enabled by local wired or wireless computer network connections. The search engine server 112 depicted in FIGURE 1 also may also operate in a distributed computing environment, which can comprise several computer systems that are interconnected via communication links, e.g., using one or more computer networks or direct connections. However, it will be appreciated by those of ordinary skill in the art that the server 112 could equally operate in a computer system having fewer or greater number of components than are illustrated in FIGURE 1. Thus, the depiction of the operating environment in FIGURE 1 should be taken as exemplary and not limiting the scope of the claims that follow.

FIGURE 2 is a block diagram depicting an automated relevance optimization system 200 for implementing an embodiment of the present invention. In one suitable implementation, the automated relevance optimization system 200 enables a search engine operator to advantageously automate the optimization of search result relevance. The automated relevance optimization system 200 includes an automated relevance optimizer process 202 that operates to collect performance data, diagnose the performance of search results based on the data, and to automatically adjust the operation of the search engine system as needed to optimize the relevance of the search results 110 that are displayed in the search Web page 106.

In one embodiment, the performance data is collected from various sources, as shown in FIGURE 2, including, among others, implicit relevance data 204 captured from the user's interaction with the search results 110, e.g., click-through data; explicit relevance data 206 collected when the user makes use of search engine help or support features provided by the search engine operator, or responds to user satisfaction surveys; human-judged test data 208 generated by human judges that periodically evaluates the relevance of search results; relevance verification test data 210 generated from verification tests that verify the validity of search results produced before and after modifications to the search relevance schema 114, i.e., to ensure that adjustments to increase the relevance of search results for one search term have not inadvertently reduced the relevance of search results for another; and sample A/B test data 212 generated from tests on groups of users, e.g., Group A and Group B, to try out adjustments to the operation of the search engine and their affect on the relevance of the

search results including the relevance of newly inserted results for which expected performance has not yet been determined, temporary modifications to the search schema to produce new and/or changed results, and other types of temporary modifications to the operation of the search engine.

5   In one embodiment, the implicit relevance data 204 is measured by the result's click-through rate (CTR), which is determined by comparing the number of times the result is displayed to the number of times a user clicks on the result after it is displayed, i.e., dividing the number of impressions by the number of clicks. The implicit relevance data 204 may also include other data tracked by the search engine server 112, such as the location of the

10   search result 110 when it was displayed on the search Web page 106, and other characteristics of the result that may influence performance, such as the color, size, font, animation, graphics, and adjacent search result performance data. The search engine server 112 is further configured to detect and filter out fraudulent clicks, as is known in the art, such as spam clicking, simulated clicks by robots, and other suspect clicks such as

15   multiple clicks from the same IP address within a certain amount of time or from unidentified sources.

          In a preferred embodiment, the implicit relevance data 204 includes data that represents complex user interaction with the search results 108 beyond the result's CTR, since the more complex user interactions are generally better predictors of relevance than the

20   CTR alone. Examples of complex user interactions include, among others, taken alone or in combination, the length of time a result was browsed, whether the result was edited, e-mailed to another party, printed, bookmarked, or whether all or portions of it was cut-and-pasted or otherwise copied for inclusion in the user's other documents.

          In one embodiment, the implicit relevance data 204 may be collected for a single

25   interaction, aggregated across all interactions in a given user session of interacting with search results, or further aggregated across all users' interactions with the same search results during their own similar searches. For example, a basic unit of implicit relevance data 204 is an atomic measurement of one user, one query, and one result interaction. There may be several atomic measurements for a given session, such as printing a result, bookmarking the

30   result in a favorites folder, sending the result in an e-mail to some friends, etc. By

aggregating the atomic measurements for all users and all queries, the implicit relevance data 204 is a sufficiently large and detailed corpus of data that is an excellent predictor of relevance. In a preferred embodiment, the automated relevance optimization system 200 collects implicit relevance data 204, as described in detail in commonly assigned copending

5 United States Patent Application No. _____, which is herein incorporated by reference. Other methods of collecting implicit relevance data 204 may be implemented without departing from the scope of the claims that follow, as long as the data is sufficiently large and detailed to be highly predictive of search result relevance.

In one embodiment, the automated relevance optimizer 202 obtains the expected

10 relevance data 214 for the results and compares the data to the actual performance, as reflected in the various sources of performance data 204, 206, 208, 210, and 212. Should the automated relevance optimizer 202 determine that the actual performance falls short of the expected performance, then an action 216 is automatically taken to adjust the operation of the search engine to increase the relevance of results and, therefore, to better their

15 performance. In one embodiment, the automated relevance optimizer 202 determines that the actual performance falls short of the expected performance only after the actual performance falls substantially below a tolerable threshold level of performance for a duration of time. This approach avoids unnecessary adjustments when the actual performance of a result happens to be erratic.

20 In one embodiment, the action 216 includes, among others, changes to the relevance schema database 114 to change the search algorithm so that the next time a search is invoked for the search term in question, a different (and improved) set of search results is produced as compared to results produced in previous searches. For example, a particular result appearing in the original set of results may be removed or re-ranked to appear further down

25 the list of results so that a new result may be inserted in its place, all in accordance with the updated search algorithm as reflected in the changed relevance schema database 114. In another embodiment, the actions 216 might include changes to the search results that are applied in real time. For example, the automated relevance optimizer 202 may intercept the search results 110 produced by the search engine 112 in accordance with the existing

30 relevance schema 114 and apply an adjustment 216 before the search results are conveyed to

the user device 102, 104 based on an up-to-the-minute diagnosis of poor performance of the results when previously displayed to other users during the last several hours.

In one embodiment, the action 216 includes, among others, changes to the expected relevance data 214 so that, the next time a search is invoked for the search term, the expected values used to measure the performance of the results are modified to better reflect the performance that the search engine operator expects before making other modifications to other aspects of the operation of the search engine.

Of course, it is understood that a combination of several of the above-described actions 216 may be taken without departing from the scope of the claims that follow. For example, the automated relevance optimizer 202 may automatically insert a new result in the search results 110 of a poor-performing search, while at the same time automatically changing the expected relevance data 214 to a lower value for an initial period of time in order to test the performance of the new result. Once a more definitive performance expectation can be ascertained, then the expected relevance data 214 may be changed accordingly.

FIGURE 3 is a block diagram depicting in further detail an arrangement of certain exemplary computing components of the search engine server 112 that are responsible for the operation of the automated relevance optimization system 200 shown in FIGURE 2. Specifically, the search engine server 112 is shown including an operating system 302, processor 306, and memory 308 to implement executable program instructions for the general administration and operation of the search engine server 112. The search engine server 112 further includes a network interface 304 to communicate with a network, such as the Internet 116, to respond to user search terms 108 and to provide search results 110. Suitable implementations for the operating system 302, processor 306, memory 308, and network interface 304 are known or commercially available, and are readily implemented by persons having ordinary skill in the art—particularly in light of the disclosure herein.

The memory 308 of the search engine server 112 includes computer executable program instructions comprising the automated relevance optimizer process 202. The automated relevance optimizer process 202 includes, among others, a data collection process 310, a diagnostic process 312, and an adjustment process 314. The data collection

process 310 is responsible for collecting performance data from the various sources previously described with reference to FIGURE 2, including the implicit relevance data 204, the explicit relevance data 206, the human-judged test data 208, the relevance verification test data 210, and the sample A/B test data 212. The data collection process 310 further normalizes the performance data based on the relative importance of the source from which the data originates. In a preferred embodiment, the implicit relevance data 204 is weighted most heavily, as it is generally considered to be a better and more objective predictor of relevance than the other sources of data. Nevertheless, in some embodiments, other performance data, such as human-judged test data, may override the implicit relevance data 204. The normalized performance data is combined in preparation for comparison with expected performance data 214.

In one embodiment, the diagnostic process 312 compares the combined and normalized performance data to the expected performance data 214. When the comparison indicates that the search results 110 are underperforming, the diagnostic process 312 attempts to determine the cause. For example, an underperforming result may be linked to an inoperative Web site, a web page that is no longer current, or a document that is no longer valid. In some cases, the result may be located in section of the Web page 106 or presented in such a way relative to the other results that is not as visible to the user as it could be.

An underperforming result may have more to do with the search term 108 than the search result 110. For instance, in some cases the search term may be one that is easily misspelled, or is too general to produce meaningful results. As an example, when a user enters the search term "Schwarzenegger" he or she is likely to misspell or mistype the name and get unwanted results. On the other hand, when the user enters the term "Arnold" he or she is likely to spell and type the term correctly, but the search term is so general, he or she will also get many unwanted results.

Once the cause of the underperforming result is determined, the adjustment process 314 is responsible for generating an appropriate action 216 to take. As already described, in one embodiment, the adjustment process 314 may generate an action 216 to modify the relevance schema database 114 to produce an optimized set of search results 110 during the next iteration of the search for the search term. Alternatively, the adjustment

process 314 may generate an action 216 to intercept the search results 110 before they are displayed to the user device 102, 104, and to automatically optimize the search results in real time. In still other embodiments, the adjustment process 314 may generate an action 216 to optimize the search results by changing the way the search results are presented on the Web page 106, rather than making substantive changes to the search results themselves. In still other embodiments, the adjustment process 314 may generate an action 216 to cause the search engine server 112 to automatically prompt the user to clarify the search term with suggested spellings or additional terms. It is understood that the adjustment process 314 may generate other actions 216 to adjust the operation of the search engine, and combine actions 216 without departing from the scope of the claims that follow.

FIGURE 4A illustrates a browser program 400 displaying a search Web page 106 in which is depicted an exemplary search engine user interface for implementing an embodiment of the present invention. The Web page 106 (FIGURE 1) may be generated by the search engine server 112 (FIGURES 1, 2) and delivered to the user's computing device 102, 104 via the Internet 116 (FIGURE 1). The search engine user interface displays the previously entered search terms 108 (FIGURE 1) in the text box 402, and prompts the user to refine the search with additional search terms, if desired, using the command button labeled "REFINE SEARCH" 404.

In the illustrated example in FIGURE 4A, the user has entered the search term "DOG FOOD." The search results are generated from a search of Web content 118 via the Internet 116. As is known in the art, the search engine 112 may execute program instructions that analyze the results and rank the "best" results for display to the user according to a predetermined criterion, such as which results are most relevant in accordance with the relevance schema database 114. The best results are generally displayed at the top of a set of results. In one embodiment, the search results may be displayed in different sections of a Web page 106, such as in a local section where the results are obtained from Web content 118 that has a local connection to the user that entered the search, e.g., Yellow Pages listings. It is understood that the search results may be displayed in a variety of different formats and in different locations on the Web page 106 without departing from the scope of the claims that follow.

By way of example only, three search results generated for the search term "DOG FOOD" are displayed directly beneath the redisplayed search entry box 402: (1) PETsMART® search result at reference numeral 406; (2) eBAY® result at reference numeral 408; and (3) Wild Alaska Salmon® Dog Food Supplement result at reference numeral 410. For purposes of illustration it is assumed that these three search results for the search term "DOG FOOD" have not yet been optimized in accordance with an embodiment of the invention, but are displayed in the order shown in accordance with the relevance schema database 114 as it existed at 1:00 p.m. The expected performance for each of the three results is listed in Table 1 below:

RESULT	EXPECTED CTR
PETsMART® (406)	15%
eBAY® (408)	10%
Wild Alaskan Salmon® (410)	8%

Table 1.

In one embodiment, during the operation of the automated relevance optimizer 202, the implicit relevance data 204 (FIGURE 2) is collected in the form of a CTR for each of the search results 406, 408, and 410, as aggregated for all users who have recently entered similar searches and obtained the illustrated search results for "DOG FOOD." In the illustrated example, the implicit relevance data 204 reveals that the actual performance of each of the search results 406, 408, and 410 is different than what was expected, and two other search results not appearing on the first page of the search results illustrated in FIGURE 4A actually outperformed all of them, as shown in Table 2 below:

RESULT	ACTUAL CTR
PETsMART® (406)	8%
eBAY® (408)	10%

RESULT	ACTUAL CTR
Wild Alaskan Salmon® (410)	5%
PETCO® (418, see FIGURE 4B)	9%
Amazon.com® (416, see FIGURE 4B)	20%

Table 2.

In the illustrated example, since the actual performance of two of the search results, PETsMART® 406 and Wild Alaskan Salmon® 410 have fallen short of the expected results, the automated relevance optimizer 202 diagnoses the possible causes of the underperforming results 406, 410 and takes an action 216 to automatically update the relevance schema database 114 so that the next time a search for "DOG FOOD" is entered, the user will see a different set of search results, namely the search results for Amazon.com® 416 and PETCO® 418 listed in Table 2, based on their superior performance.

FIGURE 4B illustrates the browser program 400 displaying another Web page, this time generated later at 3:00 p.m., as displayed at time box 416, after the automated relevance optimizer 202 has taken an action 216 to automatically optimize the relevance of the search results by updating the relevance schema database 114, and again in which the search results for a search term "DOG FOOD" are displayed. In the illustrated example, since the actual performance for the PETsMART® search result 406 that originally appeared at the top of the set of search results 110 at 1:00 p.m. did not justify keeping it there, the search result for PETsMART® has been eliminated, as shown in FIGURE 4B (or at least moved to a subsequent page). Likewise, since the actual performance for the Wild Alaskan Salmon® search result 410 that originally appeared in the third position on the first page of the set of search results 110 at 1:00 p.m. did not justify keeping it there, it too has also been eliminated, as shown in FIGURE 4B (or at least moved to a subsequent page). In place of the underperforming results 406, 410 the search engine server 112 has instead inserted new search results for Amazon.com® at reference numeral 416 and PETCO® at reference numeral 418, in accordance with the relevance schema database 114 as updated by action 216 since the earlier 1:00 p.m. time. The Amazon.com® result 416 now appears at the

top of the set of search results 110, and the PETCO® result 418 appears in the third position. The eBAY® result 408 remains in the second position as the relevance schema database 114 was unchanged with regard to eBAY®, since the result was meeting expected performance levels.

5   Note that in the above example, the Amazon.com® result 416 has nothing to do with actual dog food, as do the other results, such as PETCO® 416 and eBAY® 408. Thus, from a purely human-judged standpoint, the new set of search results may seem less relevant to the search term "DOG FOOD." But the high performance CTR of 20% revealed a trend that human judges may have missed, and that conventional content analysis may have missed,

10   namely, that the book may, at least temporarily, be a hot best-selling book that users are interested in purchasing as gifts during the holiday season. Thus, the highly predictive implicit relevance data 204 may outweigh the human-judged data 208, and produce an optimal set of results that is highly relevant and satisfactory for the majority of users. In one embodiment, another action 216 that the automated relevance optimizer 202 might have

15   taken instead of, or in addition to, updating the relevance schema database 114, would be to adjust the operation of the search engine 112 to prompt users that enter "DOG FOOD" as their search term, to further clarify whether they want a result for the new best-selling book titled "DOG FOOD," or whether they want the usual results to information about dog food itself, and where to buy dog food, etc. Once the popularity of the book "DOG FOOD" fades

20   away, the automated relevance optimizer 202 would likely take the opposite actions to update the schema 114 to eliminate the Amazon.com® result 416 and remove any prompting of the users to clarify their search. In this way, the search engine operator is able to automate the optimization of the search results to provide more relevant and timely results for users, and possibly even to generate more revenue for the search engine in increased traffic.

25   It is understood that the above example is presented by way of illustration only. In a preferred embodiment, the implicit relevance data 204 would reflect more complex user behavior than the CTR alone. For example, the implicit relevance data 204 would reflect the aggregation of a number of user behaviors with respect to the results in question, such as e-mailing the uniform record locator (URL) of the PETCO® result 418 to one or more

30   friends, making a purchase at the PETCO® Web site to which the result is linked, etc. Such

complex behaviors are generally considered to be more predictive of relevance than the CTR alone.

FIGURES 5A-5B are flow diagrams illustrating the logic performed in conjunction with the search engine server 112 and automated relevance optimizer 202 of FIGURES 2 and 3 for automating the optimization of search results in a search engine in accordance with an embodiment of the present invention. The search engine server 112 begins at the start block 502 and continues at processing block 504 to generate the search Web page 106 with search results 110 (FIGURE 1) in accordance with the existing relevance schema database 114. At processing block 506, the automated relevance optimizer process 202 collects the relevance performance data for the search results 110 from various sources, including the implicit relevance data 204, the explicit relevance data 206, the human-judged test data 208, the relevance verification test data 210, and the sample A/B test data 212. In one embodiment, collecting the implicit relevance data 204 may include aggregating the data across a local user's session or across multiple sessions, or even across multiple users where cross-user data is available from the search engine server 112. Alternatively, the implicit relevance data 204 may be already aggregated by the search engine server 112 for use by the automated relevance optimizer process 202. In any event, processing continues at process block 508, where the automated relevance optimizer process 202 compiles the various sources of collected relevance performance data into a measurement of the actual performance of a result or results under consideration, including normalizing the various sources of performance data in accordance with their relative importance. In one embodiment, the relative importance of the various sources is a measure of their value in predicting the relevance of search results, and that value may be predefined by the search engine operator, and further changed from time to time to aid the operator in fine-tuning the automated relevance optimization process. In a preferred embodiment, the value of the implicit relevance data 204 is likely to be higher than other sources of relevance performance data because it may be highly predictive due to the potentially large scale of data collection and rapidity with which the data may be collected.

In one embodiment, the automated relevance optimizer process 202 continues at process block 510 to obtain the expected relevance data 214 for the search result or results

under consideration. The expected relevance data 214 may be predefined by the search engine operator and changed from time to time to reflect changes in the expected performance of search results over time. In one embodiment, the expected relevance data 214 may have even been changed or otherwise updated by an action 216 generated by a

5   previous iteration of the automated relevance optimizer 202 to reflect any automated changes in expectations.

In one embodiment, the automated relevance optimizer process 202 continues at process block 512 to compare the measurement of the actual performance of a result or results under consideration to the expected performance, where the actual performance was

10   determined at process block 508, and the expected performance was determined at process block 510. When the comparison is unfavorable, e.g., the actual performance fall short or substantially short of the expected performance, the search results are underperforming, which may indicate a problem with the relevance of the search results. The automated relevance optimizer process attempts to diagnose the possible cause or causes of the problem.

15   For example, in some cases the result is obsolete and other newer results are now more relevant, as reflected, for example, in the implicit relevance data 204 collected for the results. In other cases, the search term for which the search results were generated is too broad or easily misspelled, and requires the search engine to prompt the user to clarify the terms. Numerous other diagnoses of possible causes of the problem may be made without departing

20   from the scope of the claims that follow.

In FIGURE 5B, the automated relevance optimizer 202 continues at decision block 514 to determine whether a problem with the relevance of the search results has been diagnosed. If not, the automated relevance optimizer 202 ends at termination oval 520. Otherwise, processing continues at process block 516, where the automated relevance

25   optimizer 202 determines what adjustment or corrective action 216 (FIGURE 2) to generate in an effort to address the problem. As described earlier, the action 216 may include adjustments to the search engine operation that cause the user to be prompted to clarify or narrow the search term in cases where the problem is related to the search term being too general or easily misspelled. In one embodiment, the action 216 may include a modification,

30   either temporary or permanent, to the relevance schema database 114, or even modifications

to the expected relevance data 214.   In still another embodiment, the action 216 may comprise one or more modifications to the search results 108 on the search Web page 106 in real time, where the search results 110 that were generated by the search engine 112 are intercepted and reranked, reordered, reformatted, removed, replaced with other results, or

5   otherwise modified in an effort to optimize the search result relevance for the user. Combinations of above-described actions 216 may be employed as well without departing from the scope of the claims that follow.

While the presently preferred embodiments of the invention have been illustrated and described, it will be appreciated that various changes may be made therein without departing

10   from the spirit and scope of the invention.   For example, as already described, in one embodiment of the present invention, the automated relevance optimization system process 202 and associated subprocesses for data collection 310, diagnosis 312, and adjustment 314, may be implemented in real time to allow for up-to-the-minute optimizations based on the latest performance data captured by the search engine server 112

15   and collected by the automated relevance optimizer 202.   In another embodiment, the automated relevance optimization system 200 processes may be implemented in batch mode to allow for data collection of performance data from a variety of sources, including implicit relevance data 204, explicit relevance data 206, human-judged test data, relevance verification test data 210, and sample A/B test data 212, and a combination of automated and

20   manual optimizations of search results.   In yet other embodiments, the automated optimization search result relevance system 200 may be limited in application to consideration of less than all sources of performance data, e.g., limited to the implicit relevance data 204, as well as limited in application to only certain types of actions 216, such as permanent modifications to the relevance schema database 114, real time updates to the

25   relevance schema database 114 or to the search results 110, or any combination thereof.